A brief survey of *a-priori* approaches to bias in Ensemble Numerical Prediction Systems

Malaquías Peña
IMSG at EMC/NCEP/NWS/NOAA

Acknowledgments: Y. Zhu, D. Hou, W. Kolczynski, E. Kalnay, Z. Toth

September 30, 2014

Why bias in DA and short time forecasts matter?

To mention a few reasons

- **(Re) analyses** are generated in the DA process. Forecast bias estimations often rely on them.
- **Origin** of systematic errors in long-range predictions can be tracked down to small systematic errors at short forecast lead times.
- **Attribution.** Generally it is easier to isolate the physical processes misrepresented at short lead times.

Numerical Weather Prediction System

Deterministic

Data Assimilation

Either bias-blind or bias-aware schemes

Interpolations

Model time integration

ζ, δ, θ, q, logPs, cloud condensate

Interpolations, couplings

Diagnostic variables

In particular surface variables: Pcpt, 2mT, surface winds, etc.

Interpolations (e.g. native grid to constant pressure)

Raw forecasts

http://www.emc.ncep.noaa.gov/index.php?branch=GFS

Numerical Climate Prediction System

Weakly coupled

DA scheme for each climate model component

Either bias-blind or bias-aware schemes

↓ Interpolations

Coupled model time integration

Dynamical variables

↓ Interpolations, couplings

Diagnostic surface variables

Pcpt, 2mT, surface winds etc.

↓ Interpolations

Raw forecasts

Ensembles can be formed by collecting runs previously initialized (lagged approach)

For forecast times of weeks and months into the future probabilistic forecasts are all that can be attempted

Numerical Ensemble Prediction System

Ensemble-Deterministic

Hybrid data assimilation

Ensemble of analyses. Additive and multiplicative noise, stochastic perturbations

↓ Interpolations

Multiple model integrations + stochastic perturbations

ζ, δ, θ, q, logPs, cloud condensate

↓ Interpolations, couplings

Ensemble diagnostic surface variables

Pcpt, 2mT, surface winds etc.

↓ Interpolations

Raw ensemble forecasts

Look so realistic that some people use them at face-value

Forecast bias in NWP systems: three general ways to approach it

- **Direct bias mitigation**: Get the bias and remove it from the forecast. Bias estimation is computed off-line or on-line
- **Account for model errors:** Add missing sub-grid variability to forecasts. Generally in combination with initial errors in the ensembles
- **Shadowing:** Find orbit in the model attractor such that its evolution maps back close to nature. Challenge to find a transformation from orbits of the forecast model to orbits in nature.

Approaches*

Numerical Stage	De-biasing: "get the bias and remove it"	Representing errors: "Compensate for the missing terms"
Data assimilation	Bias in observations (Wu and Derber 1998, Y. Zhu 2014) Bias in model (Dee and DaSilva 1998, Dee 2005)	Background error covariance inflation (Li et al 2009); State-space augmentation (Baek et al 2006)
Post-DA initialization	Vortex relocation (Liu,2010), Field alignment (Hoffman et al 2005, Ravela et al 2007), flux correction (Ji et al 1998)	Ensemble generation and centering (Toth and Kalnay, Buizza and Molteni, Wang and Bishop), Vortex relocation
Time integration	Flux correction, nudging (Saha 1992, DelSole et al 2007, Danforth et al 2008)	Stochastic perturbations (Buizza 1999, Hou 2006)
Diagnostic variables	Balance requirements (Klinker and Sardeshmukh 1992), Parameter estimation methods (KF, NN)	Stochastic perturbations in coupled systems
Typical evaluation measures	RMSE, AC, Bias=first moment SE	ROC, CRPSS, Talagrand diagram or reliability diagram to detect bias.

* Not comprehensive

Approaches*

Numerical Stage	Shadowing
Data assimilation	Pseudo-Orbit (Smith 1996, Judd and Smith 2001), Mapping (Toth and Peña 2006), State space augmentation (Baek et al 2006), Pseudo obs (Carrassi et al 2014)
Post-DA Initialization	Anomaly Initialization in numerical climate prediction systems (Schneider et al 1998, Kirtman et al, Magnussos et al 2012)
Time integration	
Diagnostic variables	
Typical evaluation measures	

* Not comprehensive

Bias correction in DA schemes

- Bias in the observing system
  - Derber and Wu 1998, Y. Zhu et al 2013: Adaptive methods
  - Model errors assumed negligible
- Bias in the model
  - Dee and DaSilva, 1998, Dee 2005: Two steps approach. First to compute bias; second to perform bias removal and obtain analysis
  - Assumption: There exists a subset of the observing system which bias is negligible compared to forecast bias

Operational DA systems generally contain both approaches

Dee and DaSilva bias correction

Analysis equation: $$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{K}[\mathbf{y}^o - H(\bar{\mathbf{x}}^f)],$$

K is the Kalman gain matrix. Function of P^f, R^{-1}

First Step: Compute the bias from previous cycle

$$\mathbf{b}^a = \mathbf{b}^f - \mathbf{K}_b[\mathbf{y}^o - H(\mathbf{x}^f - \mathbf{b}^f)]$$

$$\mathbf{K}_b = \mathbf{P}^f_{bb}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^f_{bb}\mathbf{H}^{\mathrm{T}} + \mathbf{H}\mathbf{P}^f_{xx}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}, \qquad \mathbf{P}^f_{bb} = \alpha\mathbf{P}^f_{xx}.$$

Second Step: Remove the bias

$$\mathbf{x}^a = (\mathbf{x}^f - \mathbf{b}^a) + \mathbf{K}_x[\mathbf{y}^o - H(\mathbf{x}^f - \mathbf{b}^a)]$$

Accounting for model errors in an ensemble-based DA scheme

1. Covariance inflation

$$\mathbf{P}_i^f = \mathbf{M}_{\mathbf{x}_{i-1}^a} \mathbf{P}_{i-1}^a \mathbf{M}_{\mathbf{x}_{i-1}^a}{}^T + \mathbf{Q} \qquad \text{(Ideal KF)}$$

$$\mathbf{P}_i^f = \frac{1}{k-1} \sum_{i=1}^{K} (x_i^f - \overline{x^f})(x_i^f - \overline{x^f})^T \qquad \text{(EnKF)}$$

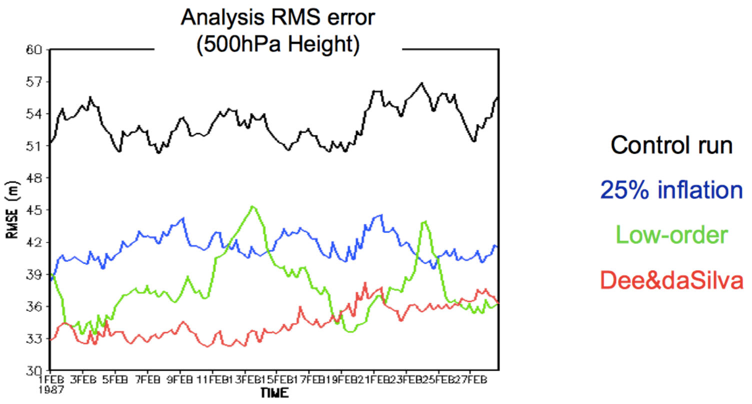
$$\widetilde{\mathbf{P}}_i^f = (1+\Delta) * \mathbf{P}_i^f = \mathbf{P}_i^f + \Delta \mathbf{P}_i^f$$

$$\rightarrow \mathbf{Q}$$

- Q represents model error
- EnKF lacks Q
- Compensates through inflation of P

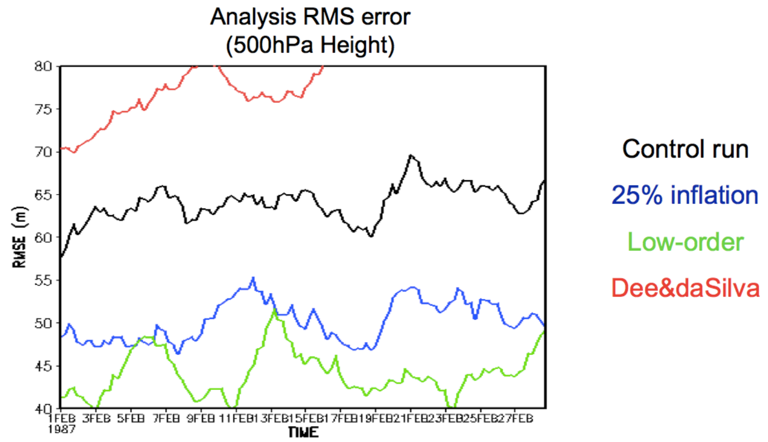
Bias correction EDAS

- Covariance inflation
  - Li et al 2009. EnKF with inflation can outperform variational DA bias correction approach in sparse data regions



Dense observation network: All schemes are better than the control run, Dee&daSilva gives best results (but it is expensive)

Sparse Observation



Sparse observation network: Dee&daSilva makes the filter divergent

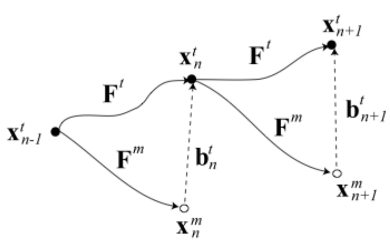
Initialization: Two general approaches

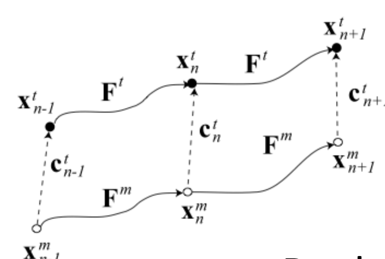
Fidelity

- Makes corrections to the initial conditions to stay as close as possible to nature
- Full field initialization
- Flux correction at initial time
- Field alignments

Mapping

- Maps nature's initial state into a state in the model climatology (attractor); returns to nature's attractor after integration.
- Used in simplified models
- Anomaly initialization





Baek et al. 2006

Initialization approaches

- Magnusson et al 2012

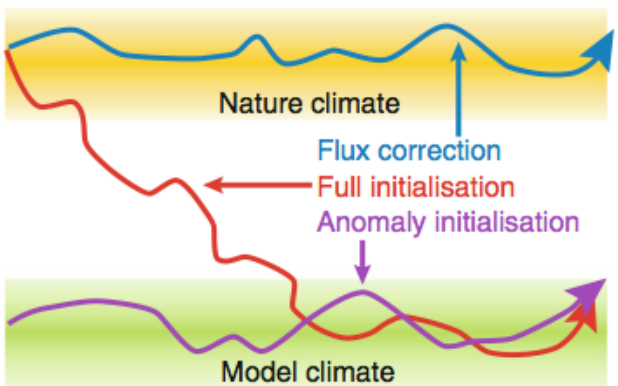
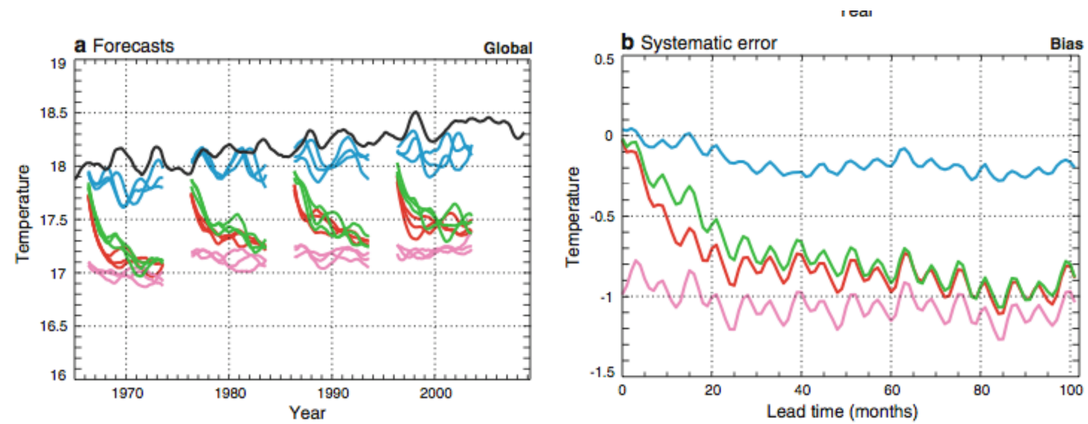


Fig. 1 Conceptual model of the forecast strategies



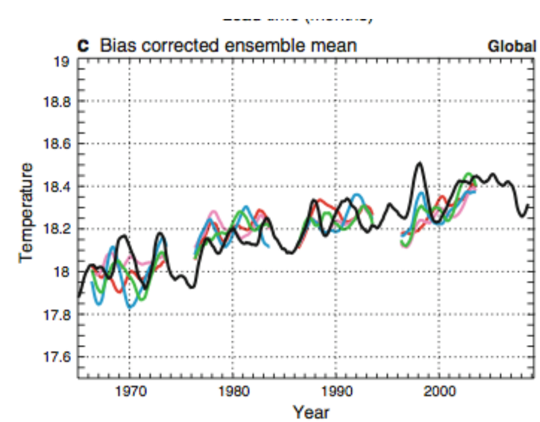


Fig. 5 Time-series of the global mean sea-surface temperature and its systematic error. *FullIni* (*red*), *Ucorr* (*green*), *UHcorr* (*blue*), *AnoIni* (*pink*) and reanalysis (*black*). 12-month running mean applied for the forecasts. The evaluation is based on 3 ensemble members

Model: ECMWF IFS (v 36r1) coupled with NEMO ocean model V3.0

Conclusions:
Best results: momentum flux correction approach
Worst: anomaly initialization

Empirical Correction Strategies

$$\underset{\textit{tendency}}{\dot{x}} = \underset{\textit{GCM}}{g(x)} + \underset{\textit{error}}{\epsilon}$$

1) Nudging based on long term biases: $\epsilon = -bias / \tau_R$

2) Relaxation: $\epsilon = (x_c - x) / \tau_R$

3) Nudging based on tendency errors: $\epsilon = \overline{\left(\frac{error}{\tau} \right)}^{t}$

From DelSole et al 2007

Empirical bias-correction

General Methdology and Idealized Studies

- Leith (1978)
- Faller and Lee (1975)
- Faller and Schemm (1977)

State-Independent Correction Improves Random Error

- Johansson and Saha (1989)
- Achatz and Branstator (1999)
- Yang and Anderson (2000)
- Danforth, Kalnay, Miyoshi (2007)

State-independent Correction Does NOT Improve Random Error

- Saha (1992)
- DelSole and Hou (1999)
- DelSole, Zhao, Dirmeyer, Kirtman (2007)

From DelSole et al 2007

Summary

1. Nudging based on tendency error clearly outperforms relaxation methods and nudging based on long-term biases.

2. Empirical correction reduces statistically significant biases in the COLAv3.2 and GFS temperature forecasts.

3. Wind biases were marginally corrected, but are small anyway.

4. Moisture biases could not be corrected significantly, but also were not amplified.

5. Empirical correction had no significant impact on random errors, or on the skill of monthly means.

6. Simple state-dependent corrections are not effective.

From DelSole et al 2007

Empirical bias correction in operational settings

- State-of-the-art models are quite complex and biases are relatively small, particularly at short lead times for some variables
  - E.g. at day 10, the SE of 500hPa forecasts is about 5% of the total RMSE
- Tuning or re-tuning of parameters has been a preferred approach. This is done constantly on each new version of the prediction system.
- Short samples (e.g. a few seasons) are used to identify bias. Longer samples are used to tune coupled numerical prediction systems

Stochastic approaches

The idea is to compensate for the "missing" terms, and, thus, increase the ensemble spread (or diversify the membership)

$$e_j(T) = e_0(0) + de_j(0) + \int_{t=0}^{T} [P_j(e_j,t) + dP_j(e_j,t) + A_j(e_j,t)]dt$$

Initial perturbations

Model Tendency

Stochastic Perturbation

Remaining tendency component (different physical parameterization or multi-model)

Buizza et al 2005

The model needs information about the sub-grid variability

Stochastic Total Tendency Perturbation

(*Hou, Toth and Zhu, 2006*)

NCEP operation – Feb. 2010

Formulation: $$\frac{\partial X_i}{\partial t} = T_i(X_i;t) + \gamma \sum_{j=1,\ldots,N} w_{i,j} T_j(X_j;t)$$

Simplification: Use finite difference form for the stochastic term

Modify the model state every 6 hours:

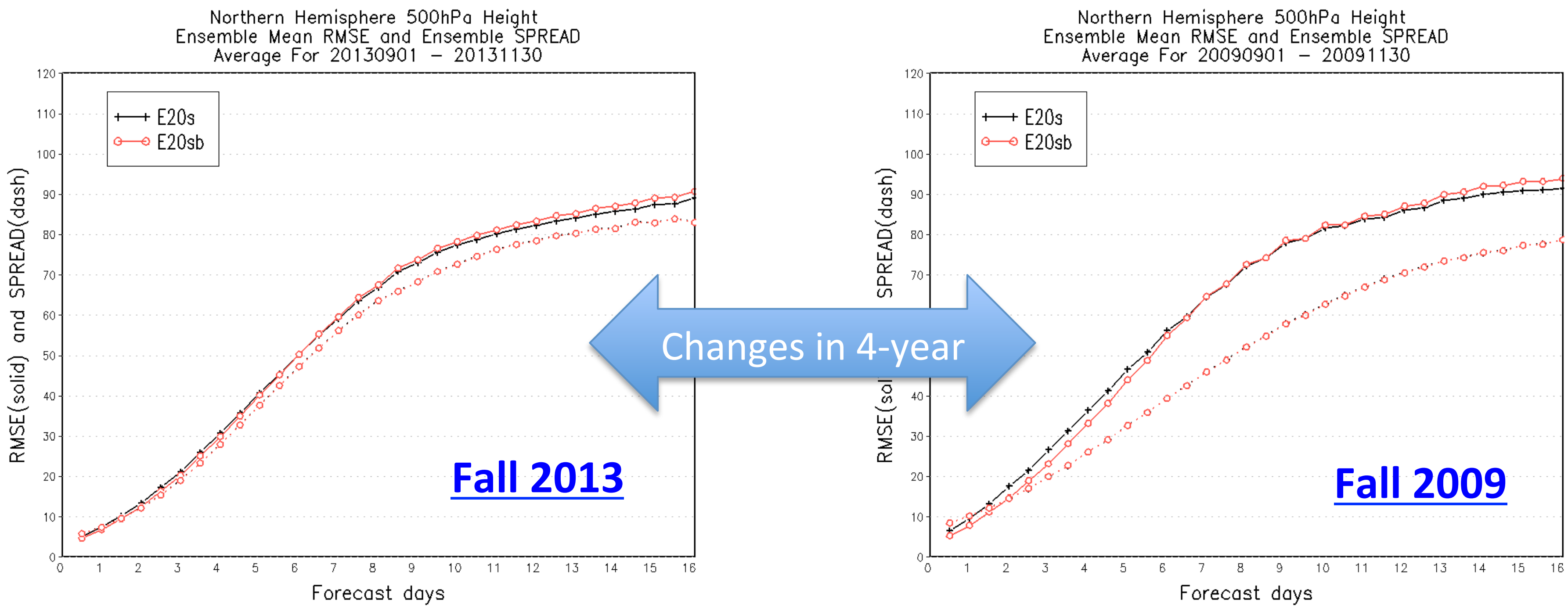
$$X_i^{'} = X_i + \gamma \sum_{j=1}^{N} w_{i,j}(t)\left\{\left[\left(X_j\right)_t - \left(X_j\right)_{t-6h}\right] - \left[\left(X_0\right)_t - \left(X_0\right)_{t-6h}\right]\right\}$$

Where w is an evolving combination matrix, and γ is a rescaling factor.

Reference:

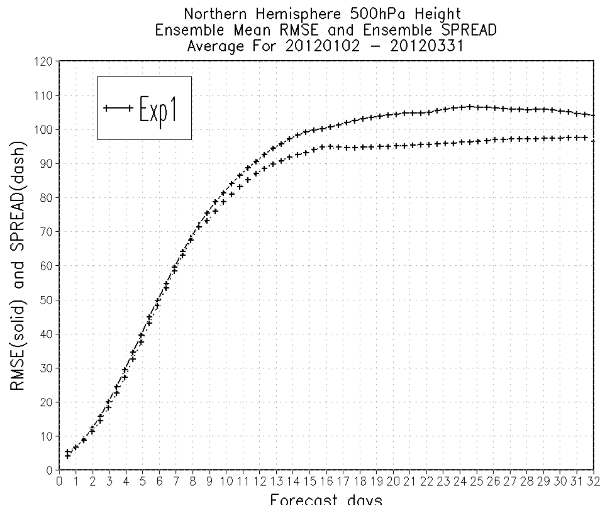
1. Hou and et al: 2008 AMS conference extended paper
2. Hou and et al: 2010 in review of Tellus

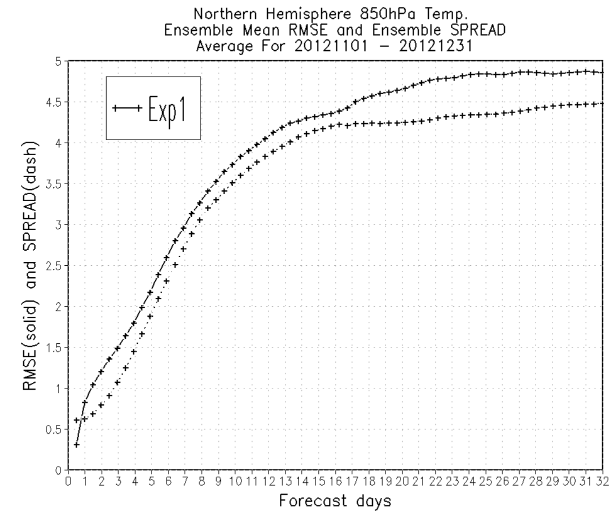
Snap shot of NCEP GEFS changes

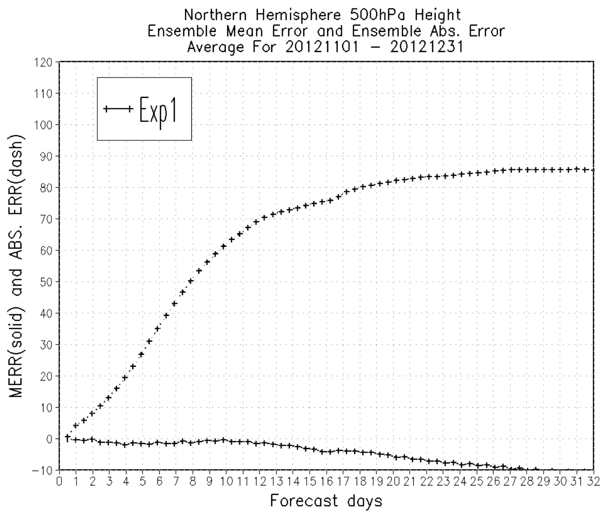


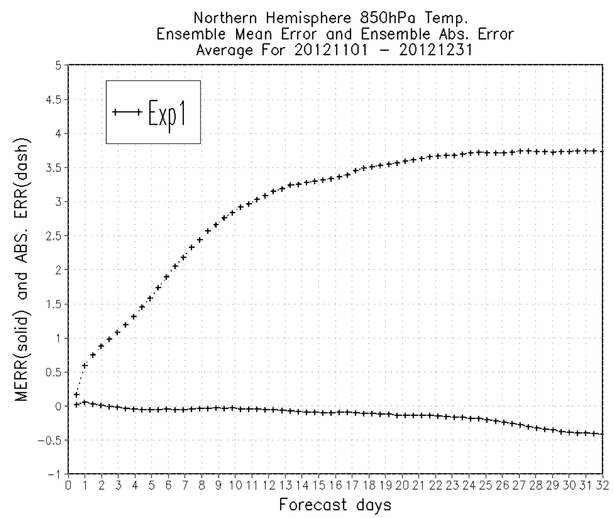
Stochastic Total Tendency Perturbation (STTP) was implemented in Feb. 2010

Results using STTP on extended GEFS



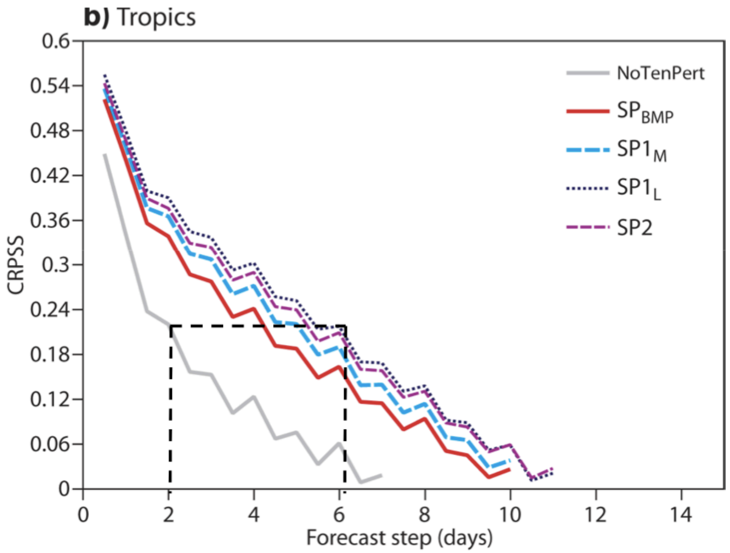




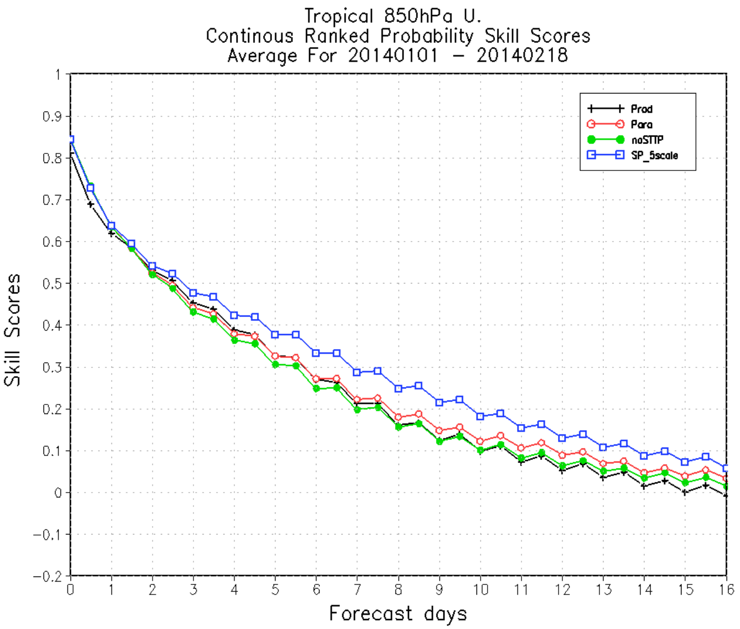


Stochastic perturbations

ECMWF T850



NCEP U850



Diagnostic variables

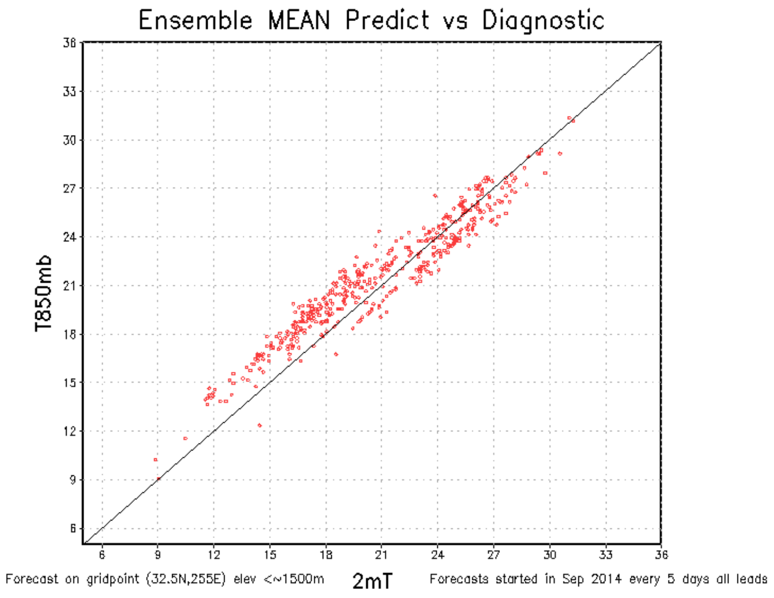
An area of much development

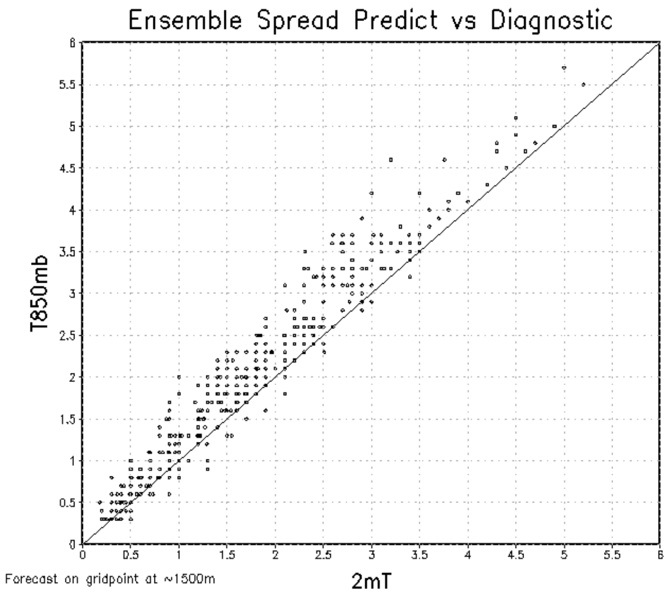
Surface variables

- Precipitation, 2mT, surface wind, etc.
- Ensemble approaches to deal with these variables quite complex.
  - Depend on surface fluxes that are difficult to measure
  - At the boundary of two climate components with two distinct time scales
  - Local feedbacks, subgrid-scale heterogeneity
- Bias correction approaches are generally *a posteriori*

Climate variables

Are not purely atmospheric but depend on other climate components





It is important to represent model uncertainty in other components of the global climate system

Stochastic perturbations vs debiasing

- Inclusion of model error schemes increases the probabilistic skill score of ensemble forecasting significantly.
- They performs best in the free atmosphere
- Multi-physics schemes increases skill at the surface. Best: Combining multi-physics schemes with parameter and stochastic perturbations
- Skill benefits comparable to calibration and/or debiasing.
- Debiasing improves reliability at the expense of resolution
- The use of model-error schemes mostly improve the reliability but at the surface there is a small but significant improvement in the resolution component.

(Berner et al 2014)

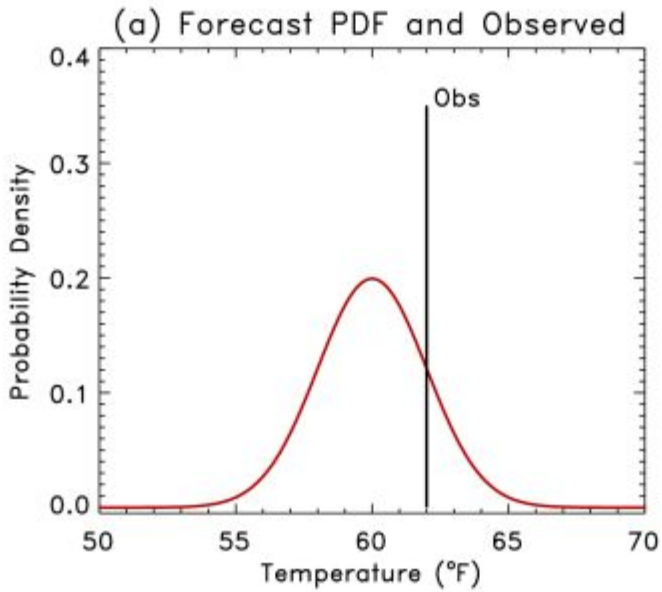
Summary

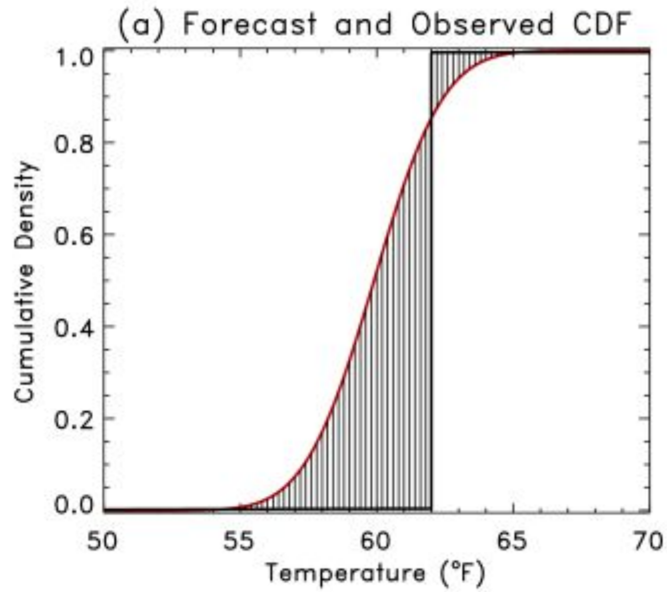
- A great deal of effort has been done to remove systematic errors *a priori*.
- Systematic error correction procedures are applied on each step of the numerical prediction system.
- New developments to represent model errors address forecast bias of the full PDF.

Thanks

Continuos Rank Probability Score

Quantifies the distance between two statistical objects, which can be two probability distributions or samples; or the distance between one point and a distribution

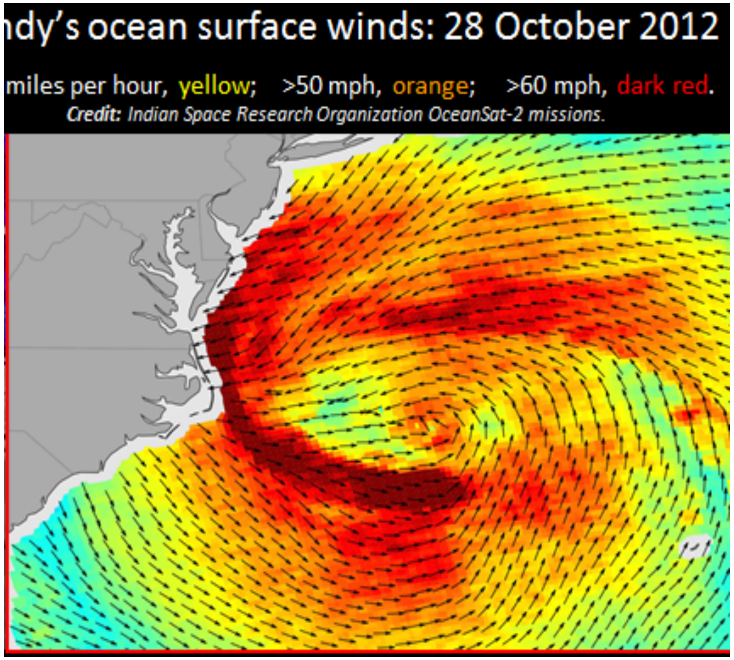




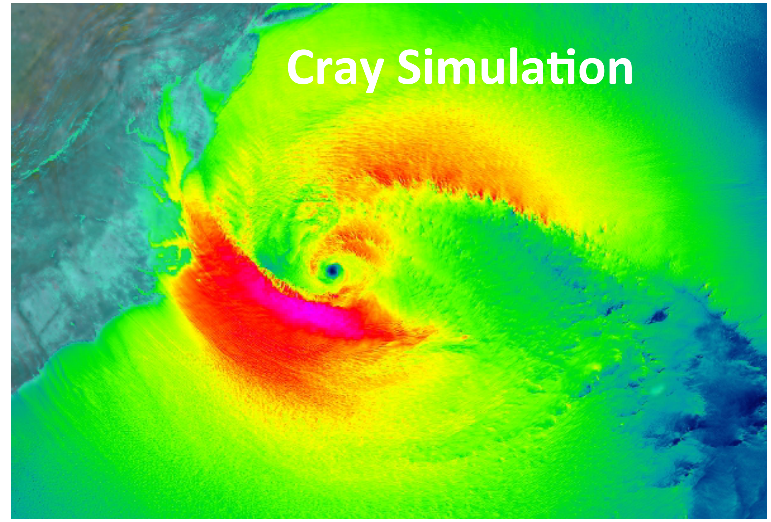
$$CRPS = \frac{1}{Cases} \sum_{i=1}^{Cases} \int_{-\infty}^{\infty} \left[F_i^{fcst}(x) - F_i^{obs}(x) \right]^2 dx$$

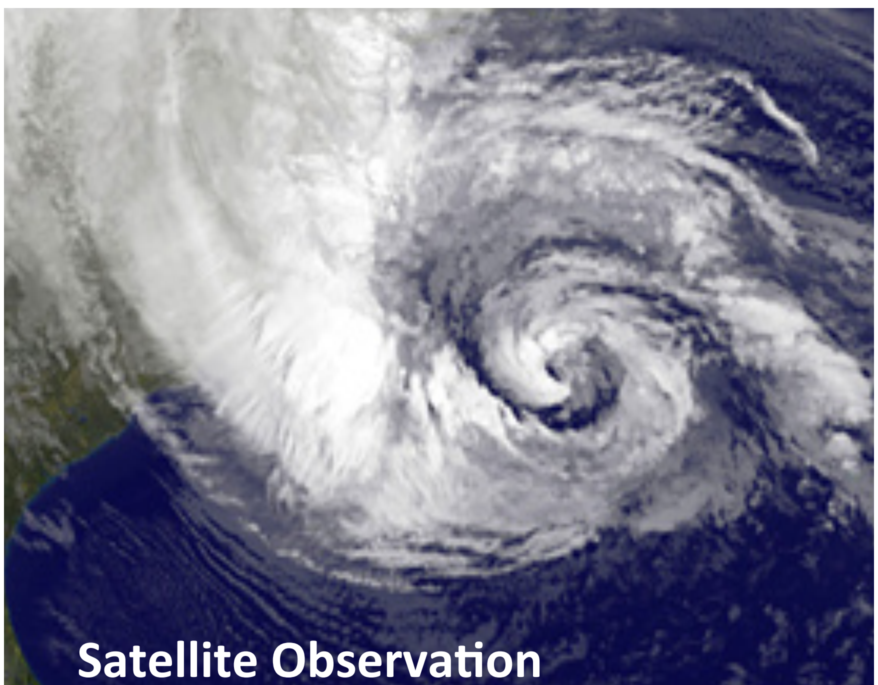
$$CRPSS = (C - CRPS) / C$$

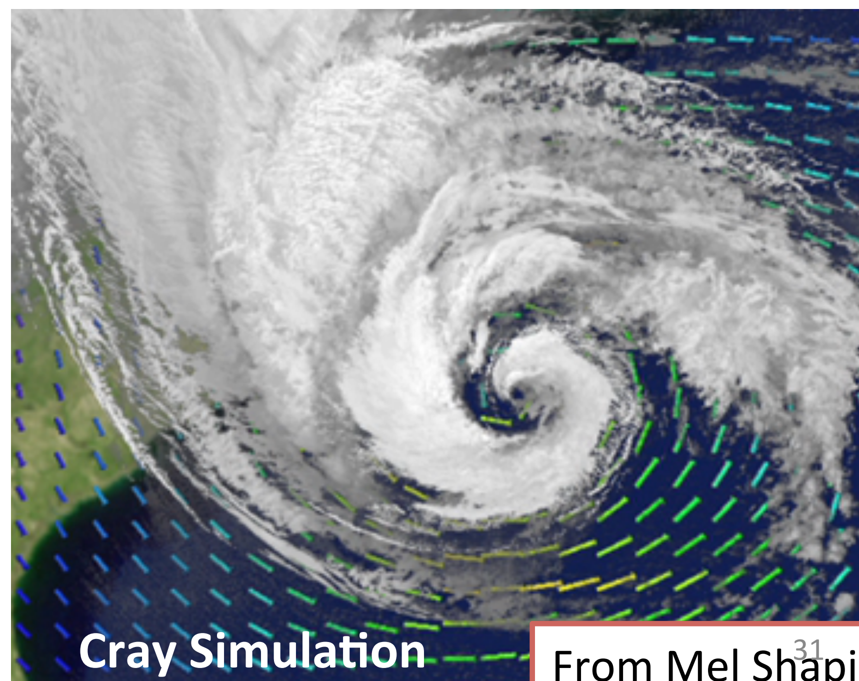
Current model outputs look very realistic



96-h Forecast







From Mel Shapiro